

# Large Language Models, Knowledge Graphs & Search Engines: A Crossroads for Answering Users' Questions

WU Viena Invited Talk

Aidan Hogan

DCC, Universidad de Chile

Instituto Milenio Fundamentos de los Datos

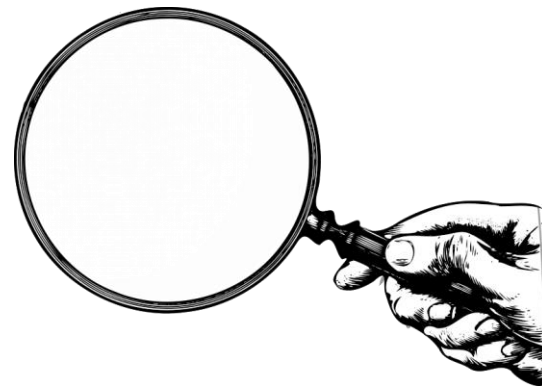


UNIVERSIDAD  
DE CHILE



Instituto Milenio  
**Fundamentos**  
de los datos

# Search Engines



### Example 1.1

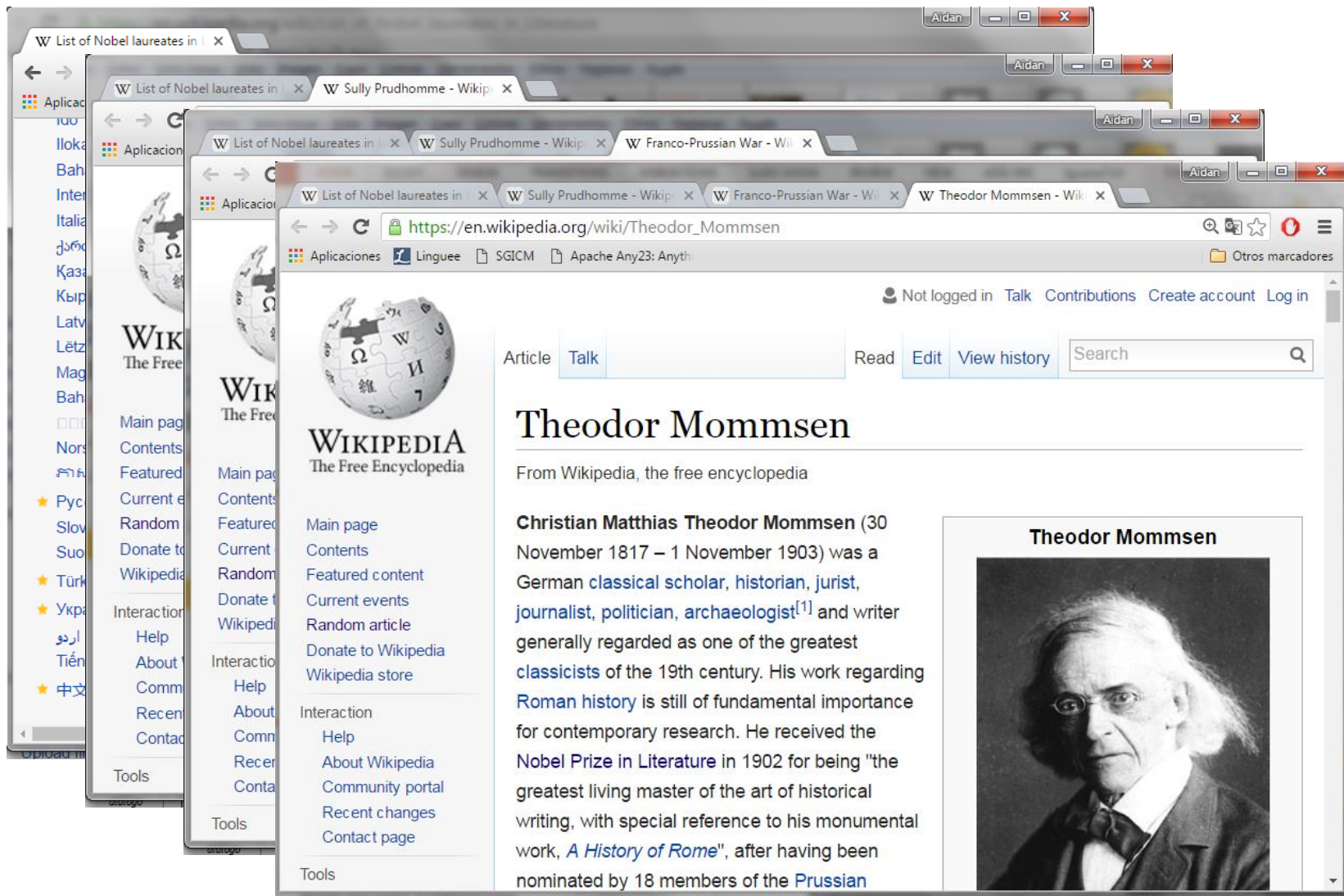


Julie is a literature student. For her thesis, she needs to find a list of Nobel Laureates in Literature who fought in a war, the year they were awarded the Nobel prize, and the name of the war(s) they fought in.



Aidan Hogan

# The Web of Data



Web  
Data



## Nobel Laureates

We list Nobel laureates ordered by category then year.

### Literature

Year	Awardee
1901	<a href="#">Sully Prudhomme</a>
...	...
1953	<a href="#">Winston Churchill</a>
1954	<a href="#">Ernest Hemingway</a>
...	...
1964	<a href="#">Jean-Paul Sartre</a> <sup>†</sup>
...	...
1983	<a href="#">William G. Golding</a>
...	...
1985	<a href="#">Claude Simon</a>
...	...

<sup>†</sup>Sartre declined.

### Peace

Year	Awardee
1901	<a href="#">Henry Dunant</a>
1901	<a href="#">Frédéric Passy</a>
...	...

## William G. Golding

Sir William Gerald Golding CBE (1911–1993) was a British novelist, playwright and poet. His most celebrated works include the novels "[Lord of the Flies](#)", and "[Rites of Passage](#)" for which he was awarded the [Booker Prize](#).

...

He joined the [Royal Navy](#) in 1940, commanding a landing ship at the [Normandy Invasion](#) on [D-Day](#).

## Invasion of Normandy

On June 6<sup>th</sup> 1944 ([D-Day](#)), the [Western Allies](#) of [World War II](#) assaulted [Nazi](#) occupation forces from the sea at [Normandy](#), as part of [Operation Overlord](#).

What Nobel Laureates fought in WWII?

William G. Golding, ...

Aidan Hogan

# The Web of Data

## Lucys Sedgyefyz

Ny sozf Lucys sedgyefyz ugwygyw ci befjyugi fhyl iyeg.

### Sofygefdgy

Iyeg	Enegwyw
1901	<a href="#">Zdssi Rgdwhummy</a>
...	...
1953	<a href="#">Nolzful Bhdgbhoss</a>
1954	<a href="#">Yglyzf Hymoljnei</a>
...	...
1964	<a href="#">Vyel-Reds Zegfgy</a> <sup>†</sup>
...	...
1983	<a href="#">Nossoem Juswolj</a>
...	...
1985	<a href="#">Bsedwy Zomul</a>
...	...

<sup>†</sup> Zegfgy wybsolyw.

### Ryeby

Iyeg	Enegwyw
1901	<a href="#">Hylgi Wdlelf</a>
1901	<a href="#">Pgýwýgob Rezzi</a>
...	...

...

## Nossoem J. Juswolj

Zog Nossoem Jygesw Juswolj BCY (1911–1993) nez e Cgofozh lutysozf, rseigojhf elw ruyf. Hoz muzf bysycgefyw nugqz olbsdwy fhy lutysz "[Sugw up fhy Psoyz](#)", elw "[Gofyz up Rezzejy](#)" pug nhobh hy nez enegwyw fhy "[Cuuqyg Rgoay](#)".

...

Hy vuolyw fhy [Guies Leti](#) ol 1940, bummelwolj e selwolj zhor ef fhy [Lugmelwi Oltezoul](#) ul [W-Wei](#).

...

## Oltezoul up Lugmelwi

Ul Vdly 6<sup>th</sup> 1944 ([W-Wei](#)), fhy [Nyzfygl Essoyz](#) up [Nugsw Neg OO](#) ezzedsfyw [Leao](#) ubbdrefoul pugbyz pgum fhy zye ef [Lugmelwi](#), ez regf up [Urygefoul Utygsugw](#).

...

Ol nhobh negz hety Lucys Sedgyefyz pudjhf?

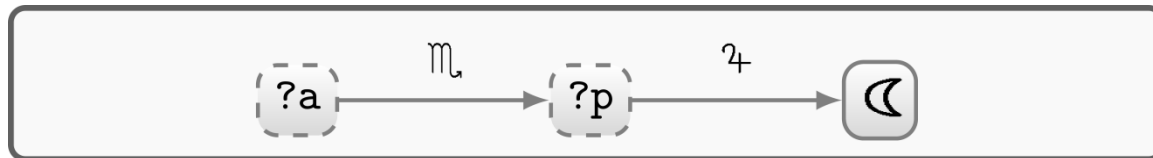
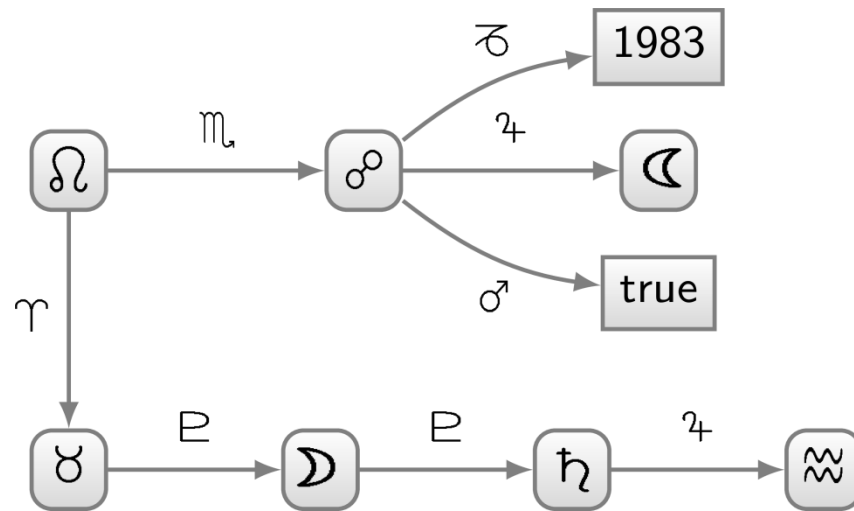
Nugsw Neg OO, ...

Aidan Hogan

# The Web of Data

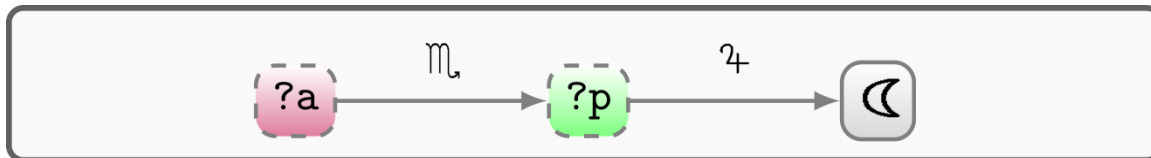
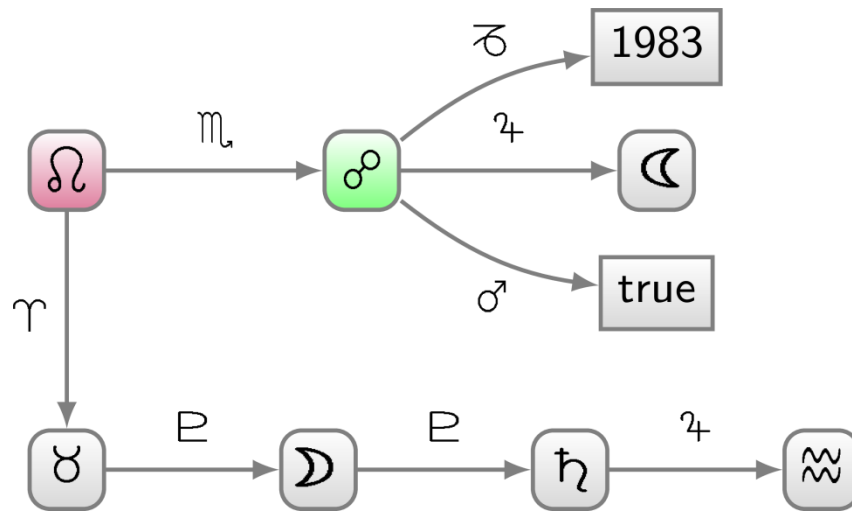
# Knowledge graphs





Aidan Hogan

# The Web of Data

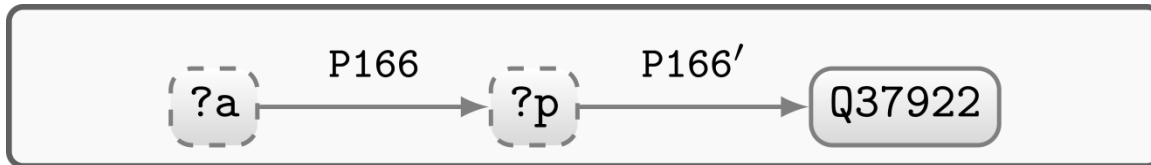
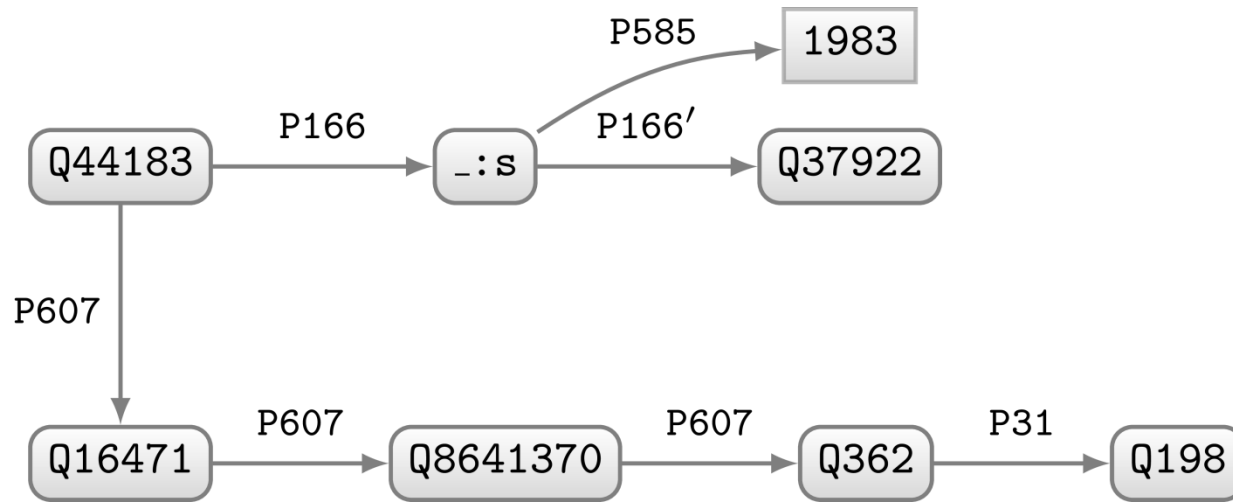


?a	?p
Ω	♂

Aidan Hogan

# The Web of Data

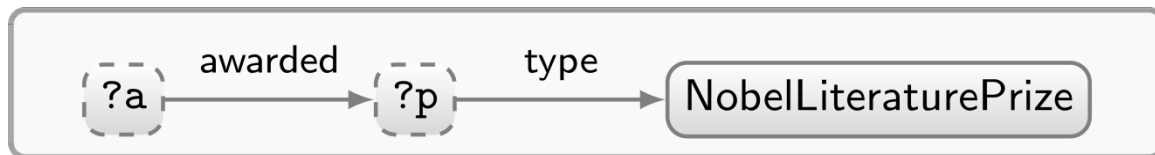
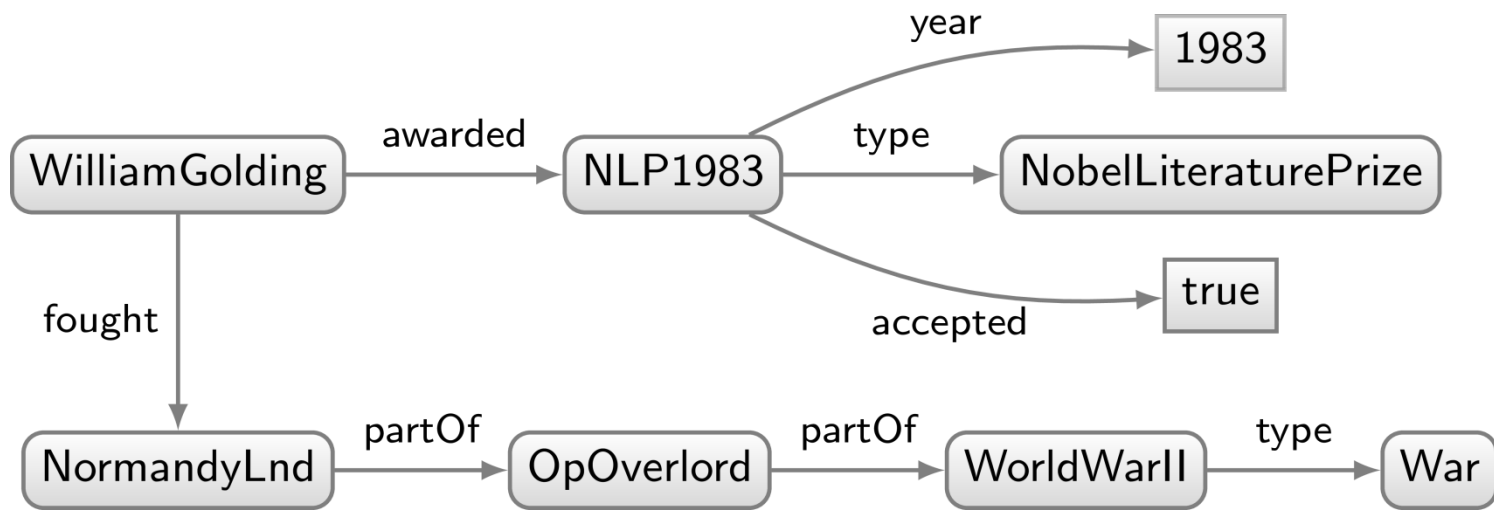




?a	?p
Q44183	_:s

Aidan Hogan

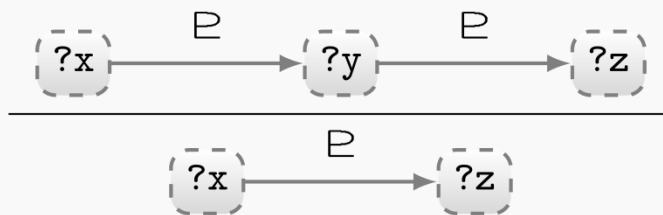
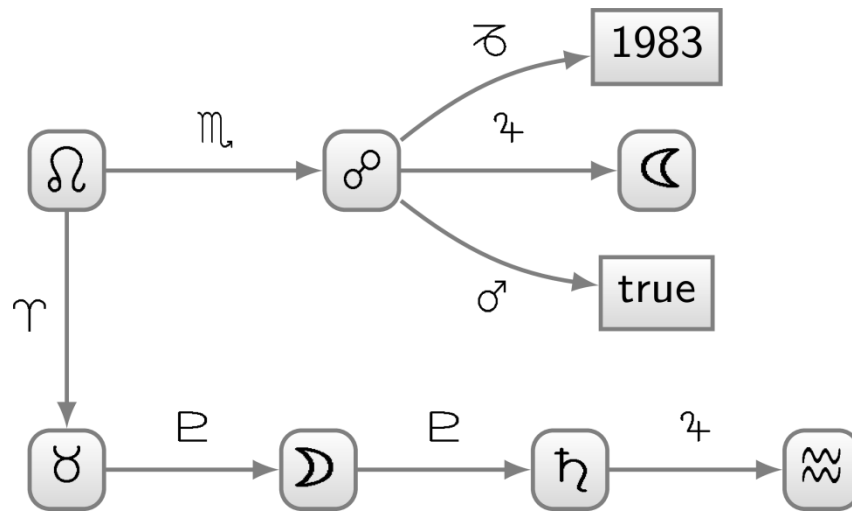
# The Web of Data



?a	?p
WilliamGolding	NLP1983

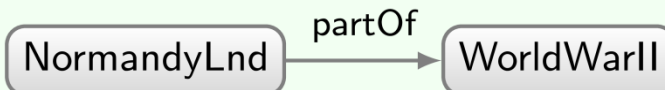
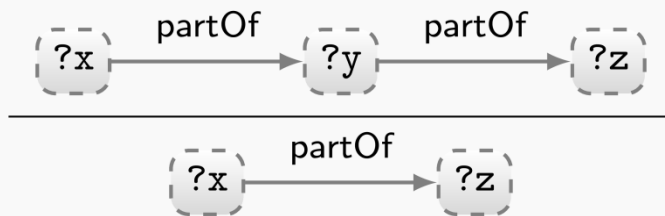
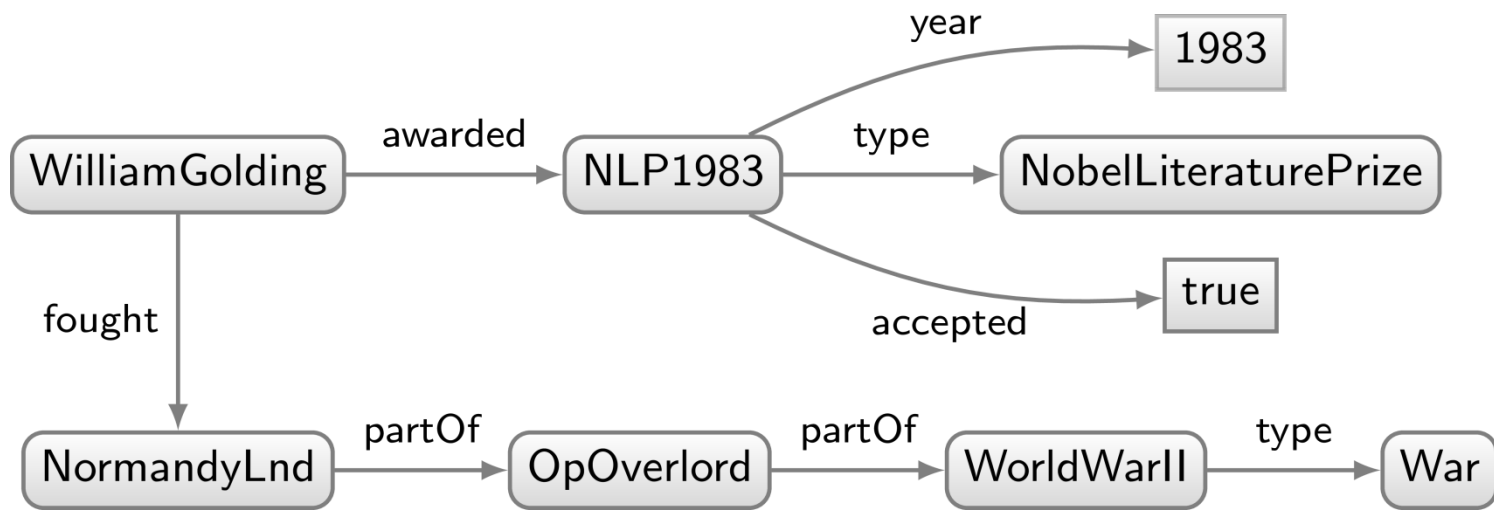
Aidan Hogan

# The Web of Data



Aidan Hogan

# The Web of Data



Aidan Hogan

# The Web of Data

# Large Language Models





## Lucys Sedgyefyz

Ny sozf Lucys sedgyefyz ugwygyw ci befjyugi fhyl iyeg.

### Sofygefdgy

Iyeg	Enegwyw
1901	<a href="#">Zdssi Rgdwhummy</a>
...	...
1953	<a href="#">Nolzful Bhdgbhoss</a>
1954	<a href="#">Yglyzf Hymoljnei</a>
...	...
1964	<a href="#">Vyel-Reds Zegfgy</a> <sup>†</sup>
...	...
1983	<a href="#">Nossoem Juswolj</a>
...	...
1985	<a href="#">Bsedwy Zomul</a>
...	...

<sup>†</sup> Zegfgy wybsolyw.

### Ryeby

Iyeg	Enegwyw
1901	<a href="#">Hylgi Wdlelf</a>
1901	<a href="#">Pgýwýgob Rezzi</a>
...	...

...

## Nossoem J. Juswolj

Zog Nossoem Jygesw Juswolj BCY (1911–1993) nez e Cgofozh lutysozf, rseigojhf elw ruyf. Hoz muzf bysycgefyw nugqz olbsdwy fhy lutysz "[Sugw up fhy Psoyz](#)", elw "[Gofyz up Rezzejy](#)" pug nhobh hy nez enegwyw fhy "[Cuuqyg Rgoay](#)".

...

Hy vuolyw fhy [Guies Leti](#) ol 1940, bummelwolj e selwolj zhor ef fhy [Lugmelwi Oltezoul](#) ul [W-Wei](#).

...

## Oltezoul up Lugmelwi

Ul Vdly 6<sup>th</sup> 1944 ([W-Wei](#)), fhy [Nyzfygl Essoyz](#) up [Nugsw Neg OO](#) ezzedsfyw [Leao](#) ubbdrefoul pugbyz pgum fhy zye ef [Lugmelwi](#), ez regf up [Urygefoul Utygsugw](#).

...

Ol nhobh negz hety Lucys Sedgyefyz pudjhf?

Nugsw Neg OO, ...

Aidan Hogan

# The Web of Data

Information may be split over multiple clauses that use references such as pronouns that may be difficult to resolve; for example, compare the “*she*” in both:

*“Jane tried to follow her late grandmother’s recipe, but **she** had not included details of the cooking times.”*

and

*“Jane tried to follow her late grandmother’s recipe, but **she** had forgotten to buy lemons”.*

Aidan Hogan

# The Web of Data



A

Jane tried to follow her late grandmother's recipe, but she had not included details of the cooking times. Who forgot to include the cooking times: Jane or her grandmother?



Jane's late grandmother forgot to include the cooking times in the recipe.

A

Jane tried to follow her late grandmother's recipe, but she had forgotten to buy lemons. Who forgot to buy lemons: Jane or her grandmother?



Jane forgot to buy lemons.

A

Why was Jane's grandmother late?



The information provided in the previous questions does not mention why Jane's grandmother was late. If you could provide more context or details, I'd be happy to help you with a specific answer.



A



Why was Jane's grandmother late?

The phrase "late grandmother" typically means that Jane's grandmother has passed away, rather than being late in the sense of tardiness. In this context, "late" is used to indicate that the person is deceased. Therefore, the sentence doesn't provide information about Jane's grandmother being late for an event; it indicates that she is no longer alive.

# Language Models sounds the Death Knell of Knowledge Graphs

Kunal Suri  
Optum, India  
kunal\_suri@optum.com

Atul Singh  
Optum, India  
atul\_singh18@optum.com

Prakhar Mishra  
Optum, India  
prakhar\_mishra29@optum.com

Swapna Sourav Rout  
Optum, India  
rout.swapnasourav@optum.com

Rajesh Sabapathy  
Optum, India  
rajesh\_sabapathy@uhc.com

**Abstract**—Healthcare domain generates a lot of unstructured and semi-structured text. Natural Language processing (NLP) has been used extensively to process this data. Deep Learning based NLP especially Large Language Models (LLMs) such as BERT have found broad acceptance and are used extensively for many applications. A Language Model is a probability distribution over a word sequence. Self-supervised Learning on a large corpus of data automatically generates deep learning-based language models. BioBERT and Med-BERT are language models pre-trained for the healthcare domain. Healthcare uses typical NLP tasks such as question answering, information extraction, named entity recognition, and search to simplify and improve processes. However, to ensure robust application of the results, NLP practitioners need to normalize and standardize them. One of the main ways of achieving normalization and standardization is the use of Knowledge Graphs. A Knowledge Graph captures concepts and their relationships for a specific domain, but their creation is time-consuming and requires manual intervention from domain experts, which can prove expensive. SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms), Unified Medical Language System (UMLS), and Gene Ontology (GO) are popular ontologies from the healthcare domain. SNOMED CT and UMLS capture concepts such as disease,

relationships among different entities makes knowledge graphs redundant for many applications.

This paper uses similar terms from SNOMED-CT KG and passes them through a language model for the healthcare domain BioRedditBERT to get a 768-dimensional dense vector representation. The paper presents the results for analyzing these embeddings. The experiments presented in the paper validate that similar terms cluster together. The paper uses simple heuristics to assign names to clusters. The results show that the cluster names match the names in the KG. Finally, the experiments demonstrate that the cosine similarity of vector representation of similar terms is high and vice versa.

Our contributions include: (i) We propose a study to demonstrate the value and application of Large Language Models (LLMs) in comparison to Knowledge Graph-based approaches for the task of synonym extraction. (ii) We extensively evaluate our approach on a standard, widely accepted dataset, and the results are encouraging.

## THE NEW YORK TIMES, TUESDAY, JULY 8, 1958.

### NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo  
of Computer Designed to  
Read and Grow Wiser

WASHINGTON, July 7 (UPI)—The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

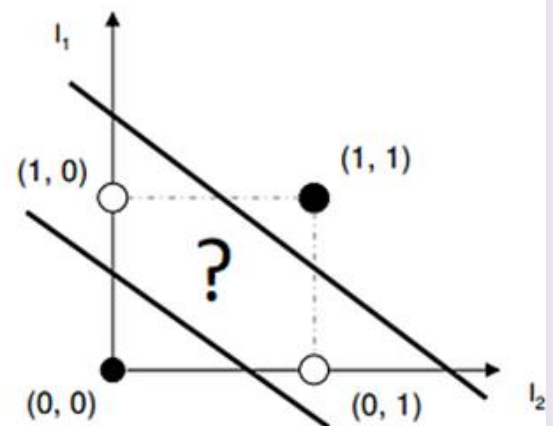
The embryo—the Weather Bureau's \$2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of \$100,000.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

and write. It is expected to be finished in about a year at a cost of \$100,000.

XOR		
$I_1$	$I_2$	out
0	0	0
0	1	1
1	0	1
1	1	0



Source: Solving XOR with a single Perceptron  
Lucas Araújo.

- Frank Rosenblatt: The Perceptron: A Method for Solving the Problems of Storage and Organization of Information (1958) 386–408 (1958)

Welcome to

EEEEEE	LL	IIII	ZZZZZZZ	AAAAA
EE	LL	II	ZZ	AA AA
EEEEE	LL	II	ZZZ	AAAAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LLLLL	IIII	ZZZZZZZ	AA AA

Eliza is a mock Rogerian psychotherapist.

The original program was described by Joseph Weizenbaum in 1966.

This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?

YOU: Men are all alike.

ELIZA: What is the connection, do you suppose ?

YOU: They're always bugging us about something or other.

ELIZA: Can you think of a specific example ?

YOU: Well, my boyfriend made me come here.

ELIZA: Is it important to you that your boyfriend made you come here ?

YOU: He says I'm depressed much of the time.

ELIZA: I am sorry to hear that you are depressed.

YOU: It's true. I am unhappy.

ELIZA: Can you explain what made you unhappy ?

YOU: ■



CULTURE DESK

# THE SAMANTHA TEST

By Brian Christian

December 30, 2013

The original program was described by Joseph Weizenbaum in 1966.

This implementa

Eliza was a milestone in computer understanding of natural language. Yet Weizenbaum was more concerned with how users seemed to form an emotional relationship with the program, which consisted of nothing more than a few hundred lines of code. “I was startled to see how quickly and how very deeply people conversing with DOCTOR became emotionally involved with the computer and how unequivocally they anthropomorphized it,” he wrote. “Once my secretary, who had watched me work on the program for many months and therefore surely knew it to be merely a computer program, started conversing with it. After only a few interchanges with it, she asked me to leave the room.” He continued, “What I had not realized is that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people.”

ELIZA: Is somethi

YOU: Men are al

ELIZA: What is th

YOU: They're al

ELIZA: Can you th

YOU: Well, my b


ELIZA: Is it impo

YOU: He says I'

ELIZA: I am sorry

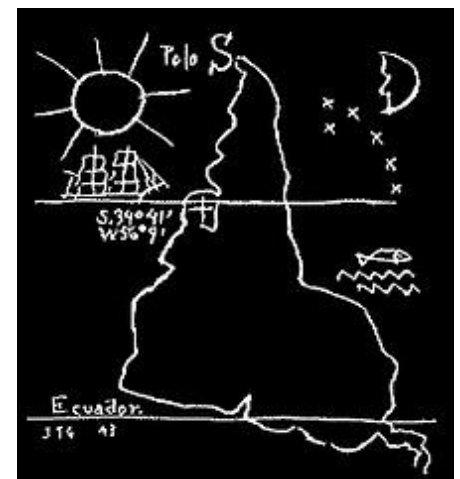
YOU: It's true.

ELIZA: Can you ex

YOU: 



# Turing tours the South





which turing award winners were born in latin america



All

Images

News

Videos

Books

Finance

Web

Tools



Wikipedia

[https://en.wikipedia.org/wiki/Turing\\_Award](https://en.wikipedia.org/wiki/Turing_Award)

## Turing Award

Only three women have been awarded the prize: Frances Allen (in 2006), Barbara Liskov (in 2008), and Shafi Goldwasser (in 2012). As of 2024, 77 people have been ...

Raj Reddy

Avi Wigderson

Robert Metcalfe

Shafi Goldwasser



Wikipedia

[https://en.wikipedia.org/wiki/Raj\\_Reddy](https://en.wikipedia.org/wiki/Raj_Reddy)

## Raj Reddy

Dabbala Rajagopal "Raj" Reddy (born 13 June 1937) is an Indian-American computer scientist and a winner of the Turing Award. He is one of the early pioneers ...

Early life and education

Career

AI Research

Awards and honors



A.M. Turing Award

[https://amturing.acm.org/reddy\\_9634208](https://amturing.acm.org/reddy_9634208)

## Raj Reddy - A.M. Turing Award Laureate

Dabbala Rajagopal (Raj) Reddy was born on June 13, 1937 in Katoor, Andhra Pradesh, India. His father, Srdenivasulu Reddy, was an agricultural landlord and ...



Britannica

[https://www.britannica.com/.../Sociology\\_Society](https://www.britannica.com/.../Sociology_Society)

## Turing Award | Definition, Winners, Prize Money, & Facts

6 days ago — The first recipient of the award, in 1966, was Alan J. Perlis, an American computer scientist who wrote the compiler for the ALGOL computer ...



Medium · Oksana Pochapska

A



Which Turing Award winners were born in Latin America?

As of my last update in September 2021, Adolfo Guzmán-Arenas is a Turing Award winner born in Latin America. He was born in Guatemala. He received the Turing Award in 2002 along with his colleague Manuel Blum for their work on computational complexity theory and the development of the concept of a "proof system."

Please note that my information might not be current, so I recommend verifying with the latest sources to confirm the most up-to-date information about Turing Award winners from Latin America.

A



Which Turing Award winners were born in Latin America?

As of my last update in September 2021, there are no Turing Award winners who were born in Latin America. The Turing Award is often considered the "Nobel Prize of Computer Science" and is awarded annually by the Association for Computing Machinery (ACM). The recipients have primarily been from the United States, Europe, and Asia, reflecting in part the historical investment and focus on computer science research in these regions.

However, it's important to note that the field of computer science is increasingly global, and outstanding contributions are being made by researchers from all around the world, including Latin America. While there have not been Turing Award winners born in Latin America as of my last update, this could change in the future as the field continues to evolve and diversify.

Please verify this information for the most current data, as it might have changed after my last training data in September 2021.

A



Can you give me a SPARQL query for Wikidata to find Turing Award winners who were born in Latin America?

Certainly!

```
# Turing Award winners born in Latin America
SELECT ?person ?personLabel ?birthPlace ?birthPlaceLabel ?award ?awardLabel WHERE {
  ?person p:P166 ?statement.
  ?statement ps:P166 wd:Q7251; pq:P582 ?awardDate. # Turing Award
  ?person wdt:P19 ?birthPlace. # Place of birth
  ?birthPlace wdt:P30 wd:Q12521. # Latin America
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
}
ORDER BY ?awardDate
```

# SEs vs. KGs vs. LLMs

**IL  
BUONO** **IL  
BRUTTO** **IL  
CATTIVO**

**SEARCH ENGINES** ASSUME BOTH HUMANS AND MACHINES ARE STUPID

∴ WIDELY USED AND WIDELY USEFUL



**KNOWLEDGE GRAPHS** ASSUME MACHINES ARE STUPID AND HUMANS SMART

∴ BARELY USED AND BARELY USEFUL



**LARGE LANGUAGE MODELS** ASSUME MACHINES ARE SMART AND HUMANS STUPID

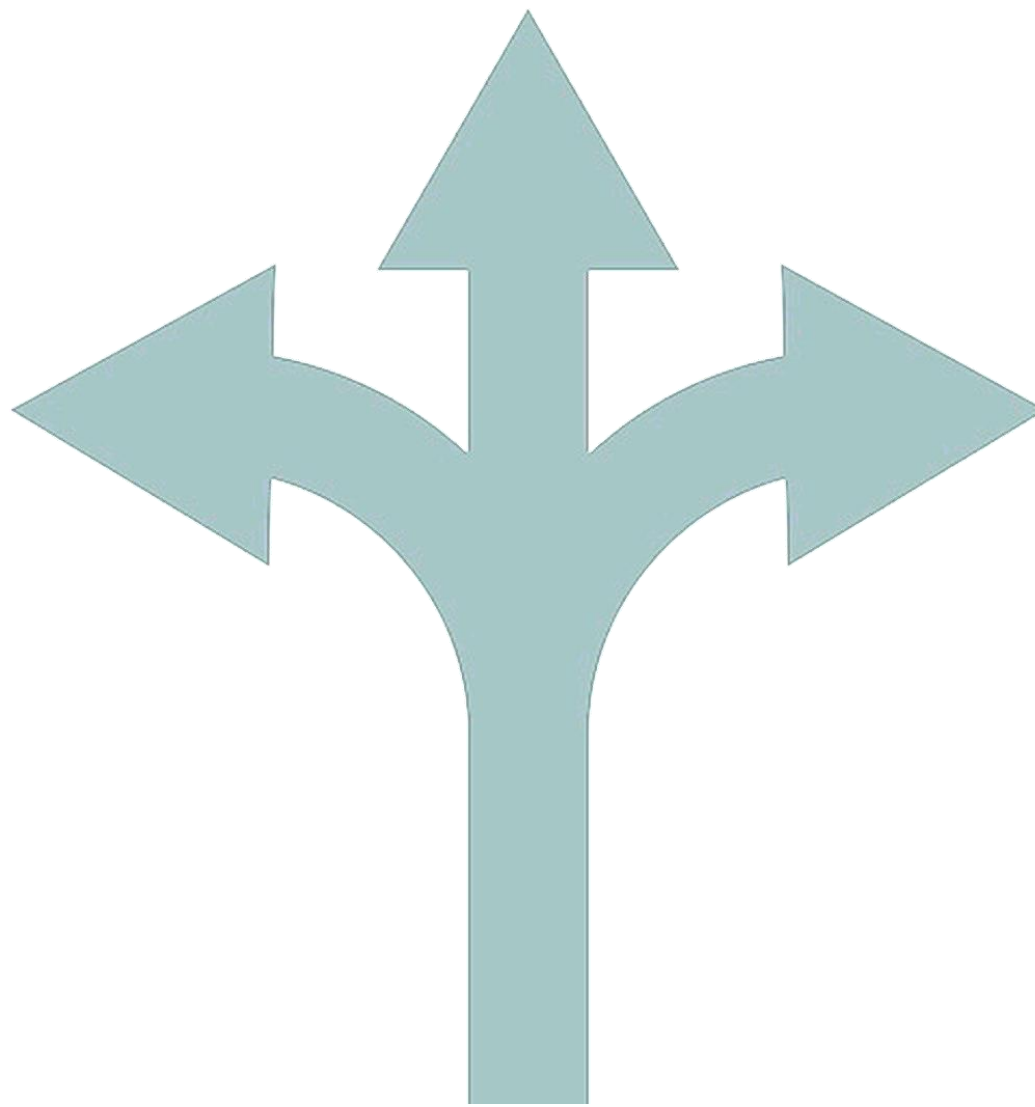
∴ WIDELY USED AND BARELY USEFUL





Dimension	SE	KG	LM
Precision	+ stores corpus – noisy content	+ stores corpus + precise operators	– abstracts corpus – hallucinations
Coverage	+ broad coverage	+ domain specific – patchy coverage	+ broad coverage – poor long tail
Freshness	+ quick updates + news often text	+ deprecation – structure lag	– slow updates – cold start
Generation	– no generation	+ ontologies/rule + graph learning	+ text generation
Synthesis	– no integration – no synthesis	+ data integration + synthesis	+ text integration + synthesis
Transparency	+ provenance – opaque ranks	+ algorithmic + provenance	– black box – no provenance
Determinism	+ deterministic	+ deterministic	– unstable results – randomness
Curation	+ curatable – opaque ranks	+ curatable	– indirect curation – unpredictable
Fairness	+ content as-is – bias in ranks	+ content as-is – bias in coverage	– generative biases – needs safeguards
Usability	+ natural language + simple queries	– structured – complex queries	+ natural language + conversational
Expressivity	– simple queries – ambiguity	+ complex queries – lacks nuance	+ complex queries – ambiguity
Efficiency	+ retrieval-based + simple queries	+ retrieval-based – complex queries	– inference-based – costly training
Multilingual	– lang. dependent	+ lang. agnostic – manual labels	+ multilinguality – variable results
Context	– limited context – not interactive	– limited context – not personalizable	+ in-context learn. + personalizable







Category	Subcategory	Example	SE	KG	LM
Facts	POPULAR	<i>Who directed the movie Spotlight?</i>	+ good coverage − noisy content	+ high precision − poor coverage	+ good coverage − noisy content
	LONG-TAIL	<i>Which galaxy is closest to the Sunflower Galaxy?</i>	+ good coverage − needle in haystack	+ high precision − sparse coverage	− limited storage − hallucinations
	MULTI-HOP	<i>Which Turing Award winners were born in Latin America?</i>	− no reasoning − single-shot search	+ formal reasoning + structured queries	+ latent reasoning − hallucinations
	ANALYTICAL	<i>How many U.S. Congress Members are younger than 50?</i>	− no datatypes − no aggregation	+ rich datatypes + aggregation	− no datatypes − no aggregation
Explanations	COMMONSENSE	<i>How do snakes move?</i>	+ good coverage + text output	− poor coverage − structured output	+ good coverage + text output
	CAUSAL	<i>What caused the dancing plague of 1518?</i>	+ good coverage + text output	+ long tail − structured output	+ good coverage + text output
	EXPLORATORY	<i>Who was Williamina Fleming?</i>	+ text output + ranked results	+ graph algorithms + browsing + navigation	+ interactive + synthesis
Planning	INSTRUCTIVE	<i>How do I tie a Windsor Knot?</i>	+ multimedia + diverse results	− poor coverage − non-didactic output	+ interactive − no multimedia
	RECOMMENDATION	<i>Should I pack warm clothes for Iceland in June?</i>	+ diverse results + ranked results	− poor coverage − no recommendations	+ interactive + synthesis
	SPATIO-TEMPORAL	<i>What kid-friendly Italian restaurants are near Disneyland?</i>	+ events & maps − no integration	+ integration + s.-t. operators	− lacks freshness − no s.-t. operators
Advice	LIFESTYLE	<i>How can I improve my work/life balance?</i>	+ diverse results + ranked results	− poor coverage − lacks nuance	+ interactive + synthesis
	CULTURAL	<i>Should I tip bartenders in Canada?</i>	+ diverse results + ranked results	− poor coverage − lacks nuance	+ synthesis − cultural bias
	PHILOSOPHICAL	<i>Is the death penalty ever acceptable?</i>	+ diverse results + ranked results	− poor coverage − lacks nuance	+ interactive + synthesis



Category	Subcategory	Example	SE	KG	LM
Facts	POPULAR	Who directed the movie <i>Spotlight</i> ?	+ good coverage − noisy content	+ high precision − poor coverage	+ good coverage − noisy content
	LONG-TAIL	Which galaxy is closest to the Sunflower Galaxy?	+ good coverage − needle in haystack	+ high precision − sparse coverage	− limited storage − hallucinations
	MULTI-HOP	Which Turing Award winners were born in Latin America?	− no reasoning − single-shot search	+ formal reasoning + structured queries	+ latent reasoning − hallucinations
	ANALYTICAL	How many U.S. Congress Members are younger than 50?	− no datatypes − no aggregation	+ rich datatypes + aggregation	− no datatypes − no aggregation
Explanations	COMMONSENSE	How do snakes move?	+ good coverage + text output	− poor coverage − structured output	+ good coverage + text output
	CAUSAL	What caused the dancing plague of 1518?	+ good coverage + text output	+ long tail − structured output	+ good coverage + text output
	EXPLORATORY	Who was Williamina Fleming?	+ text output + ranked results	+ graph algorithms + browsing + navigation	+ interactive + synthesis
Planning	INSTRUCTIVE	How do I tie a Windsor Knot?	+ multimedia + diverse results	− poor coverage − non-didactic output	+ interactive − no multimedia
	RECOMMENDATION	Should I pack warm clothes for Iceland in June?	+ diverse results + ranked results	− poor coverage − no recommendations	+ interactive + synthesis
	SPATIO-TEMPORAL	What kid-friendly Italian restaurants are near Disneyland?	+ events & maps − no integration	+ integration + s.-t. operators	− lacks freshness − no s.-t. operators
Advice	LIFESTYLE	How can I improve my work/life balance?	+ diverse results + ranked results	− poor coverage − lacks nuance	+ interactive + synthesis
	CULTURAL	Should I tip bartenders in Canada?	+ diverse results + ranked results	− poor coverage − lacks nuance	+ synthesis − cultural bias
	PHILOSOPHICAL	Is the death penalty ever acceptable?	+ diverse results + ranked results	− poor coverage − lacks nuance	+ interactive + synthesis

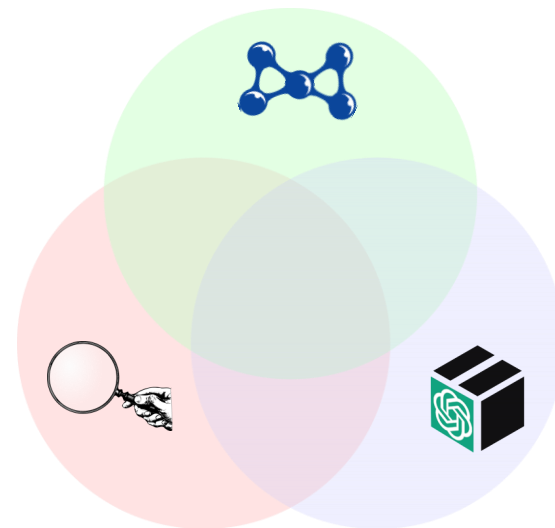


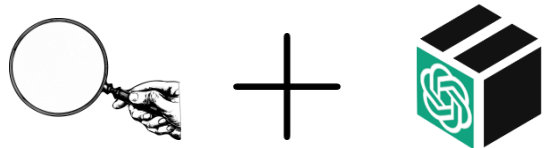
Category	Subcategory	Example	SE	KG	LM
Facts	POPULAR	<i>Who directed the movie Spotlight?</i>	+ good coverage − noisy content	+ high precision − poor coverage	+ good coverage − noisy content
	LONG-TAIL	<i>Which galaxy is closest to the Sunflower Galaxy?</i>	+ good coverage − needle in haystack	+ high precision − sparse coverage	− limited storage − hallucinations
	MULTI-HOP	<i>Which Turing Award winners were born in Latin America?</i>	− no reasoning − single-shot search	+ formal reasoning + structured queries	+ latent reasoning − hallucinations
	ANALYTICAL	<i>How many U.S. Congress Members are younger than 50?</i>	− no datatypes − no aggregation	+ rich datatypes + aggregation	− no datatypes − no aggregation
Explanations	COMMONSENSE	<i>How do snakes move?</i>	+ good coverage + text output	− poor coverage − structured output	+ good coverage + text output
	CAUSAL	<i>What caused the dancing plague of 1518?</i>	+ good coverage + text output	+ long tail − structured output	+ good coverage + text output
	EXPLORATORY	<i>Who was Williamina Fleming?</i>	+ text output + ranked results	+ graph algorithms + browsing + navigation	+ interactive + synthesis
Planning	INSTRUCTIVE	<i>How do I tie a Windsor Knot?</i>	+ multimedia + diverse results	− poor coverage − non-didactic output	+ interactive − no multimedia
	RECOMMENDATION	<i>Should I pack warm clothes for Iceland in June?</i>	+ diverse results + ranked results	− poor coverage − no recommendations	+ interactive + synthesis
	SPATIO-TEMPORAL	<i>What kid-friendly Italian restaurants are near Disneyland?</i>	+ events & maps − no integration	+ integration + s.-t. operators	− lacks freshness − no s.-t. operators
Advice	LIFESTYLE	<i>How can I improve my work/life balance?</i>	+ diverse results + ranked results	− poor coverage − lacks nuance	+ interactive + synthesis
	CULTURAL	<i>Should I tip bartenders in Canada?</i>	+ diverse results + ranked results	− poor coverage − lacks nuance	+ synthesis − cultural bias
	PHILOSOPHICAL	<i>Is the death penalty ever acceptable?</i>	+ diverse results + ranked results	− poor coverage − lacks nuance	+ interactive + synthesis



Category	Subcategory	Example	SE	KG	LM
Facts	POPULAR	<i>Who directed the movie Spotlight?</i>	+ good coverage − noisy content	+ high precision − poor coverage	+ good coverage − noisy content
	LONG-TAIL	<i>Which galaxy is closest to the Sunflower Galaxy?</i>	+ good coverage − needle in haystack	+ high precision − sparse coverage	− limited storage − hallucinations
	MULTI-HOP	<i>Which Turing Award winners were born in Latin America?</i>	− no reasoning − single-shot search	+ formal reasoning + structured queries	+ latent reasoning − hallucinations
	ANALYTICAL	<i>How many U.S. Congress Members are younger than 50?</i>	− no datatypes − no aggregation	+ rich datatypes + aggregation	− no datatypes − no aggregation
Explanations	COMMONSENSE	<i>How do snakes move?</i>	+ good coverage + text output	− poor coverage − structured output	+ good coverage + text output
	CAUSAL	<i>What caused the dancing plague of 1518?</i>	+ good coverage + text output	+ long tail − structured output	+ good coverage + text output
	EXPLORATORY	<i>Who was Williamina Fleming?</i>	+ text output + ranked results	+ graph algorithms + browsing + navigation	+ interactive + synthesis
Planning	INSTRUCTIVE	<i>How do I tie a Windsor Knot?</i>	+ multimedia + diverse results	− poor coverage − non-didactic output	+ interactive − no multimedia
	RECOMMENDATION	<i>Should I pack warm clothes for Iceland in June?</i>	+ diverse results + ranked results	− poor coverage − no recommendations	+ interactive + synthesis
	SPATIO-TEMPORAL	<i>What kid-friendly Italian restaurants are near Disneyland?</i>	+ events & maps − no integration	+ integration + s.-t. operators	− lacks freshness − no s.-t. operators
Advice	LIFESTYLE	<i>How can I improve my work/life balance?</i>	+ diverse results + ranked results	− poor coverage − lacks nuance	+ interactive + synthesis
	CULTURAL	<i>Should I tip bartenders in Canada?</i>	+ diverse results + ranked results	− poor coverage − lacks nuance	+ synthesis − cultural bias
	PHILOSOPHICAL	<i>Is the death penalty ever acceptable?</i>	+ diverse results + ranked results	− poor coverage − lacks nuance	+ interactive + synthesis

SEs + KGs + LLMs





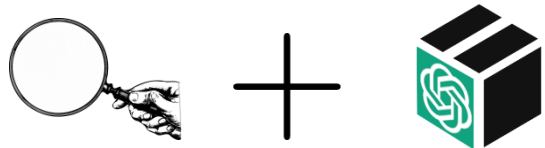
A

Which Turing Award winners were born in Latin America?



*Searched 6 sites*

As of now, the only Turing Award winner born in Latin America is Manuel Blum. He was born in Caracas, Venezuela, in 1938. Blum received the Turing Award in 1995 for his contributions to the foundations of computational complexity theory and its application to cryptography and program checking(Wikipedia,Wikipedia).



A

How many congress members are younger than 50?



*Searched 5 sites*

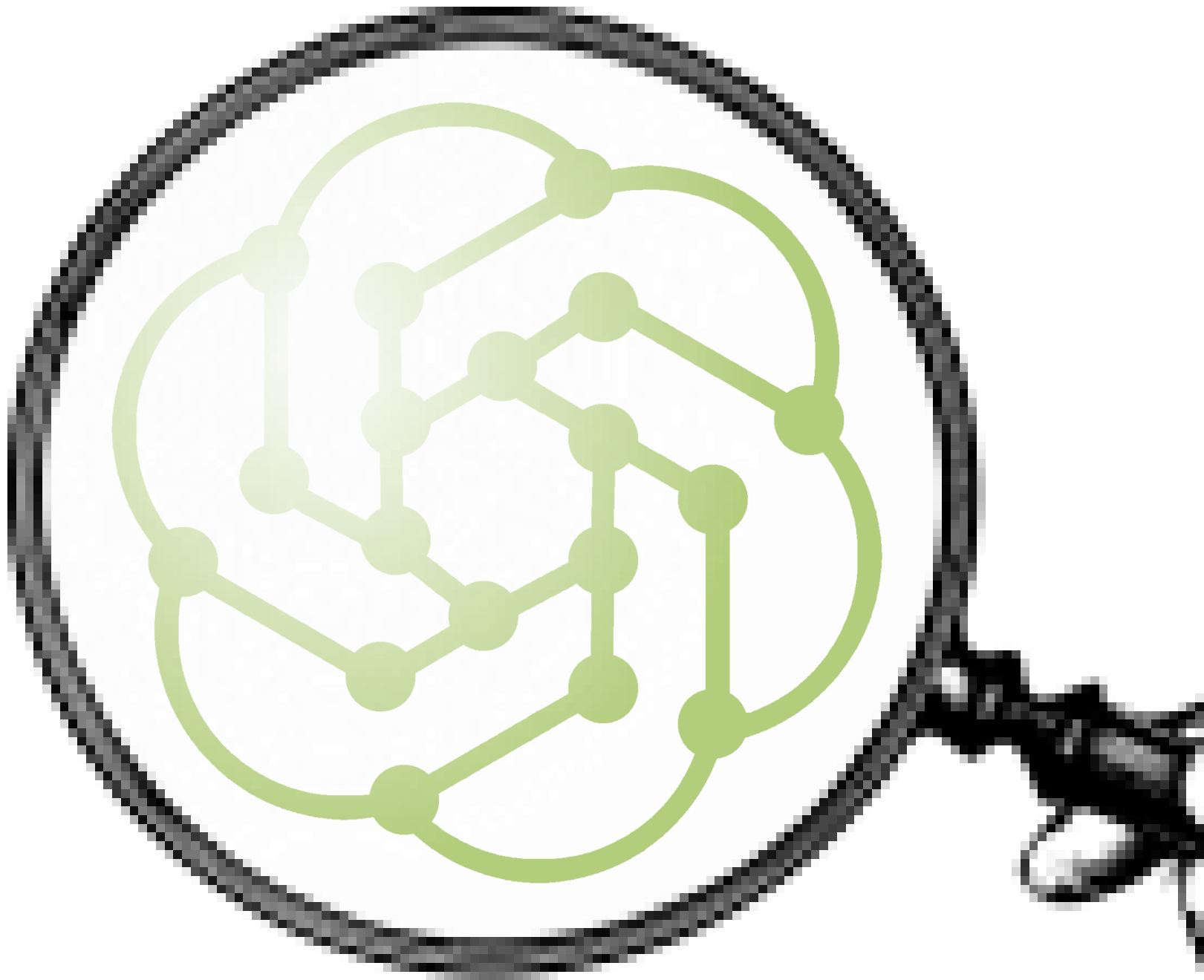
As of the 118th Congress in 2024, there are approximately 139 members of Congress who are under the age of 50. This includes members from both the House of Representatives and the Senate. In the House, the median age is 57.9, with a significant number of younger members, particularly among the freshmen representatives. Millennials (aged 27-42) and Generation X (aged 43-58) have a substantial presence, collectively making up about 46% of the total congressional membership (Pew Research Center, FiscalNote).

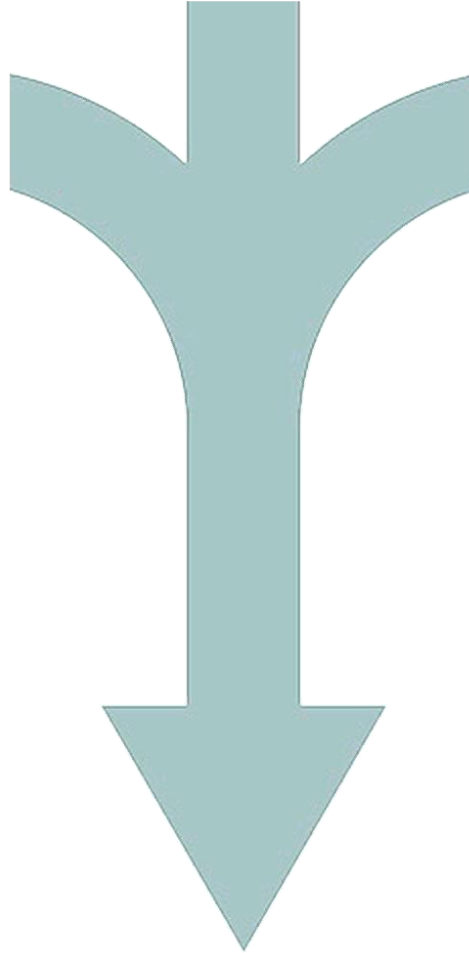




A

```
# Members of U.S. congress younger than 50
SELECT (COUNT(DISTINCT ?person) AS ?answer)
WHERE {
  ?person wdt:P31 wd:Q5 .
  ?person p:P39 ?stmt .
  { ?stmt ps:P39 wd:Q13218630 . }
  UNION { ?stmt ps:P39 wd:Q4416090 }
  MINUS { ?stmt pq:P582 ?edate }
  ?person wdt:P569 ?bdate .
  BIND(YEAR(NOW()) - YEAR(?bdate) -
        IF(MONTH(NOW())>MONTH(?bdate),0,
            IF(MONTH(NOW())=MONTH(?bdate) && DAY(NOW())>=DAY(?bdate),0,1)
        ) AS ?age)
  FILTER(?age < 50)
}
```



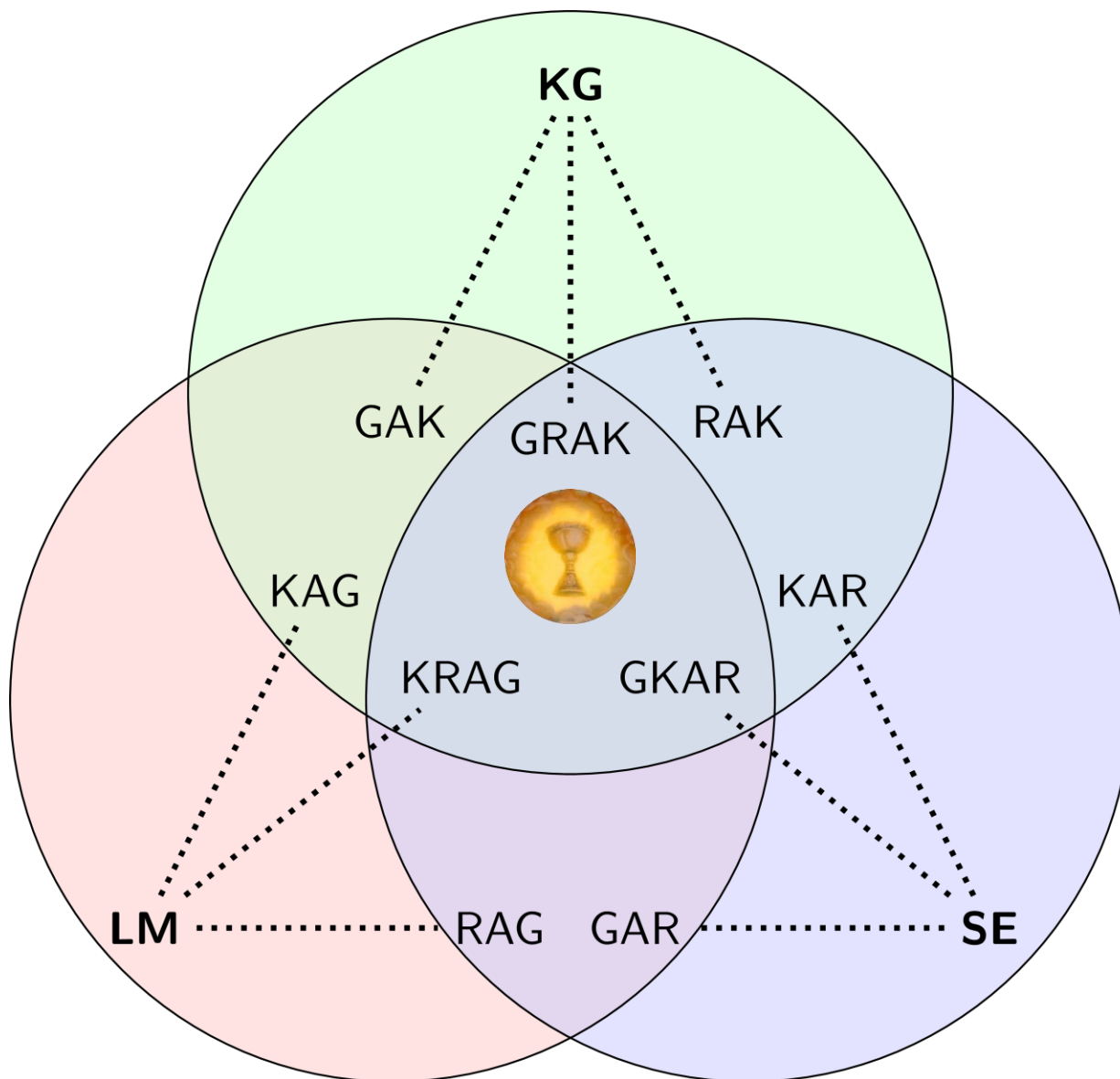


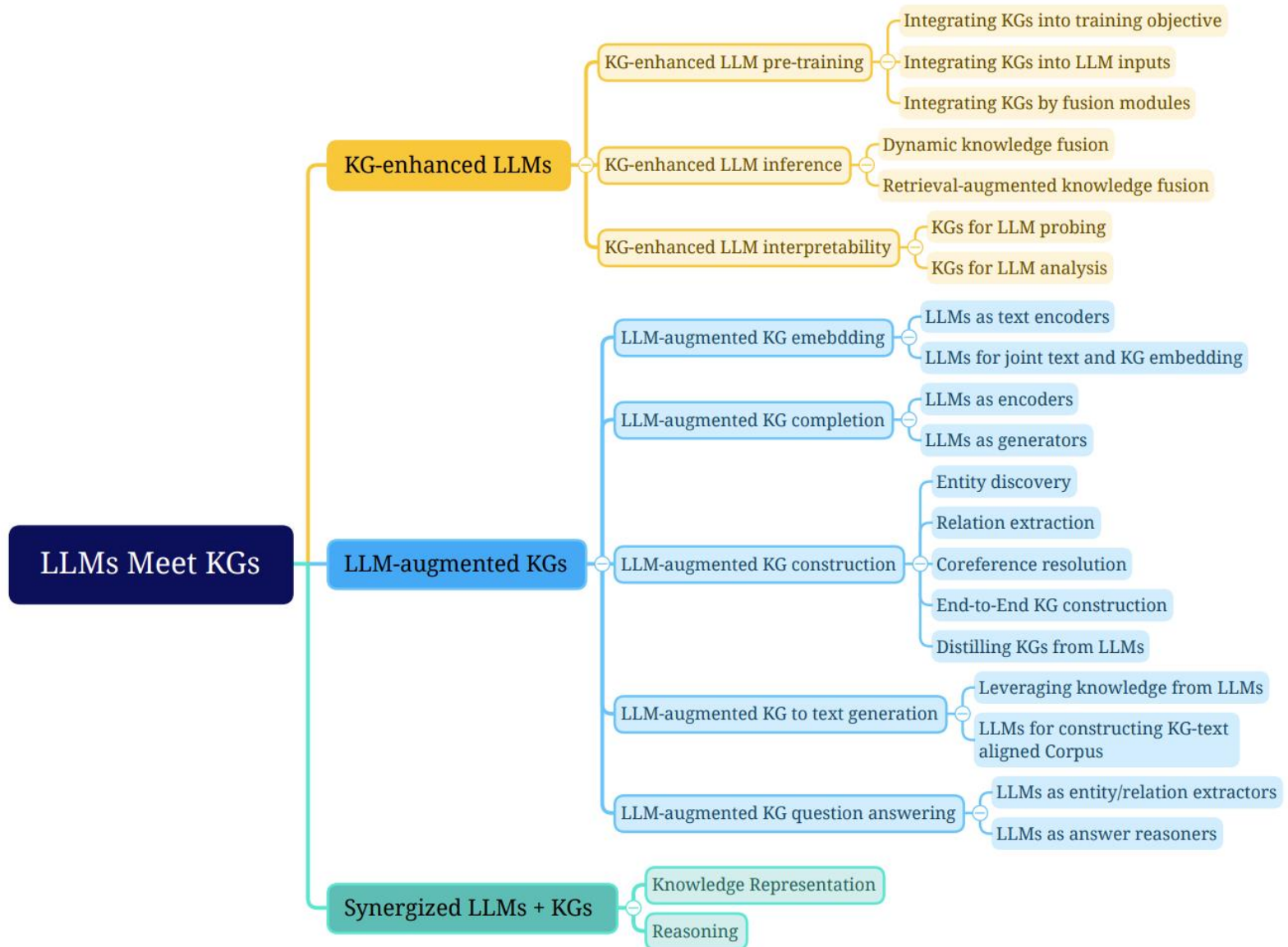


Dimension	SE	KG	LM	SE + KG + LM
Precision	+ stores corpus − noisy content	+ stores corpus − noisy content	− abstracts corpus − hallucinations	+ stores corpus + precise operators
Coverage	+ broad coverage	+ domain specific − patchy coverage	+ broad coverage − poor long tail	+ broad coverage + domain specific
Freshness	+ quick updates + news often text	+ deprecation − structure lag	− slow updates − cold start	+ quick updates + deprecation
Generation	− no generation	+ ontologies/rule + graph learning	+ text generation	+ data generation + text generation
Synthesis	− no integration − no synthesis	+ data integration + synthesis	+ text integration + synthesis	+ data integration + text integration
Transparency	+ provenance − opaque ranks	+ algorithmic + provenance	− black box − no provenance	+ algorithmic + provenance
Determinism	+ deterministic	+ deterministic	− unstable results − randomness	+ deterministic
Curation	+ curatable − opaque ranks	+ curatable	− indirect curation − unpredictable	+ curatable
Fairness	+ content as-is − bias in ranks	+ content as-is − bias in coverage	− generative biases − needs safeguards	+ content as-is
Usability	+ natural language + simple queries	− structured − complex queries	+ natural language + conversational	+ natural language + conversational
Expressivity	− simple queries − ambiguity	+ complex queries − lacks nuance	+ complex queries − ambiguity	+ complex queries
Efficiency	+ retrieval-based + simple queries	+ retrieval-based − complex queries	− inference-based − costly training	+ retrieval-based
Multilingual	− lang. dependent	+ lang. agnostic − manual labels	+ multilinguality − variable results	+ lang. dependent + multilinguality
Context	− limited context − not interactive	− limited context − not personalizable	+ in-context learn. + personalizable	+ in-context learn. + personalizable



Category	Subcategory	Example	SE	KG	LM	SE + KG + LM
Facts	POPULAR	<i>Who directed the movie Spotlight?</i>	+ good coverage − noisy content	+ high precision − poor coverage	+ good coverage − noisy content	+ good coverage + high precision
	LONG-TAIL	<i>Which galaxy is closest to the Sunflower Galaxy?</i>	+ good coverage − needle in haystack	+ high precision − sparse coverage	− limited storage − hallucinations	+ good coverage + high precision
	MULTI-HOP	<i>Which Turing Award winners were born in Latin America?</i>	− no reasoning − single-shot search	+ formal reasoning + structured queries	+ latent reasoning − hallucinations	+ formal reasoning + structured queries
	ANALYTICAL	<i>How many U.S. Congress Members are younger than 50?</i>	− no datatypes − no aggregation	+ rich datatypes + aggregation	− no datatypes − no aggregation	+ rich datatypes + aggregation
Explanations	COMMONSENSE	<i>How do snakes move?</i>	+ good coverage + text output	− poor coverage − structured output	+ good coverage + text output	+ good coverage + text output
	CAUSAL	<i>What caused the dancing plague of 1518?</i>	+ good coverage + text output	+ long tail − structured output	+ good coverage + text output	+ good coverage + text output
	EXPLORATORY	<i>Who was Williamina Fleming?</i>	+ text output + ranked results	+ graph algorithms + browsing + navigation	+ interactive + synthesis	+ hybrid output + hybrid interactivity
Planning	INSTRUCTIVE	<i>How do I tie a Windsor Knot?</i>	+ multimedia + diverse results	− poor coverage − non-didactic output	+ interactive − no multimedia	+ interactive + diverse results
	RECOMMENDATION	<i>Should I pack warm clothes for Iceland in June?</i>	+ diverse results + ranked results	− poor coverage − no recommendations	+ interactive + synthesis	+ interactive + synthesis
	SPATIO-TEMPORAL	<i>What kid-friendly Italian restaurants are near Disneyland?</i>	+ events & maps − no integration	+ integration + s.-t. operators	− lacks freshness − no s.-t. operators	+ s.-t. operators + integration
Advice	LIFESTYLE	<i>How can I improve my work/life balance?</i>	+ diverse results + ranked results	− poor coverage − lacks nuance	+ interactive + synthesis	+ diverse results + interactive
	CULTURAL	<i>Should I tip bartenders in Canada?</i>	+ diverse results + ranked results	− poor coverage − lacks nuance	+ synthesis − cultural bias	+ diverse results + synthesis
	PHILOSOPHICAL	<i>Is the death penalty ever acceptable?</i>	+ diverse results + ranked results	− poor coverage − lacks nuance	+ interactive + synthesis	+ diverse results + synthesis





- Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, Xindong Wu: Unifying Large Language Models and Knowledge Graphs: A Roadmap. CoRR abs/2306.08302 (2023).

G(R)AK



A



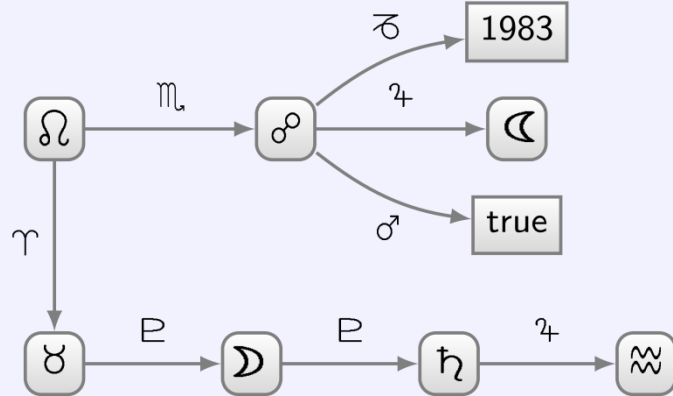
Can you give me a SPARQL query for Wikidata to find Turing Award winners who were born in Latin America?

Certainly!

```
# Turing Award winners born in Latin America
SELECT ?person ?personLabel ?birthPlace ?birthPlaceLabel ?award ?awardLabel WHERE {
  ?person p:P166 ?statement.
  ?statement ps:P166 wd:Q7251; pq:P582 ?awardDate. # Turing Award
  ?person wdt:P19 ?birthPlace. # Place of birth
  ?birthPlace wdt:P30 wd:Q12521. # Latin America
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
}
ORDER BY ?awardDate
```

Niche Problem (1)	6	26	33	35	13	$\bar{x} \approx 3.204, p \approx 0.047$
Redundant w/ML (2)	3	14	23	31	42	$\bar{x} \approx 3.841, p \approx 0.000$
Unreliable Publishers (3)	20	42	25	20	6	$\bar{x} \approx 2.558, p \approx 0.000$
Ontological Disagreement (4)	14	30	22	33	14	$\bar{x} \approx 3.027, p \approx 0.821$
Publishing Costly (5)	14	21	20	32	26	$\bar{x} \approx 3.310, p \approx 0.014$
Lacks Incentives (6)	25	49	19	14	6	$\bar{x} \approx 2.354, p \approx 0.000$
Verbose Standards (7)	22	22	21	27	19	$\bar{x} \approx 2.991, p \approx 0.946$
Won't Scale (8)	13	12	38	29	21	$\bar{x} \approx 3.292, p \approx 0.011$
Lacks Usable Tools (9)	32	51	14	11	4	$\bar{x} \approx 2.143, p \approx 0.000$
Decentralisation Costly (10)	11	17	23	25	36	$\bar{x} \approx 3.518, p \approx 0.000$
	1	2	3	4	5	

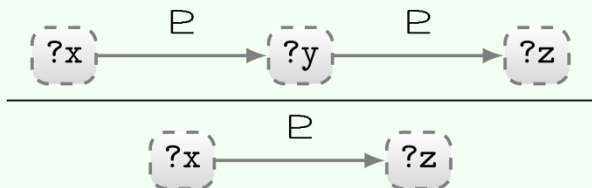
## DATA



## QUERY



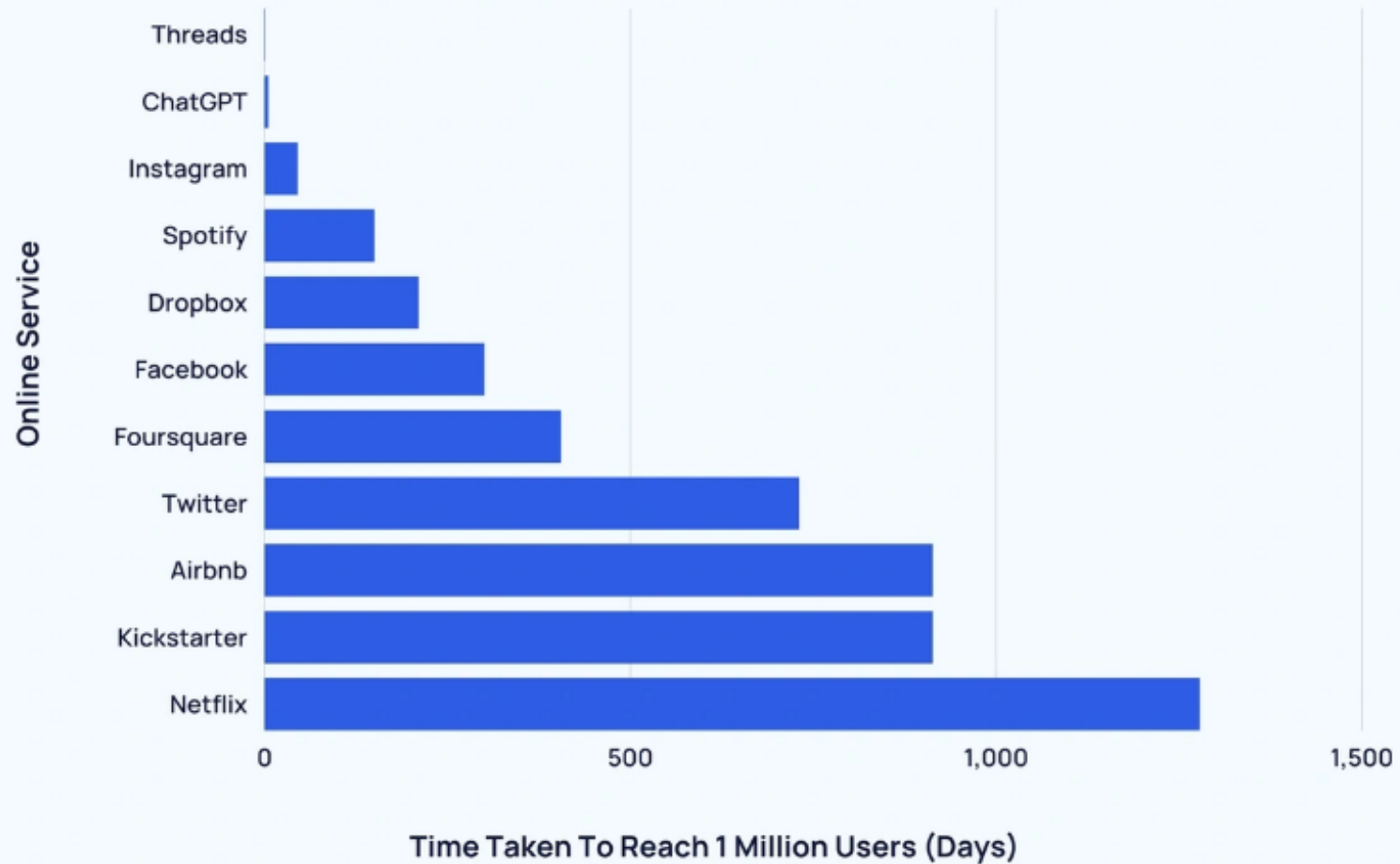
## LOGIC



## LANGUAGE



## Time taken to reach 1 million users



Wrap-up

- Neural vs. symbolic an old debate, but why not both?
- SEs, LLMs and KGs address different needs
- SEs, LLMs and KGs are complementary
- SEs + LLMs + KGs = A fertile research topic
- Not just RAG: KAG, KRAG, GAK, GRAK, etc.
- Back to basics: focus on what users need

The real power of LLMs will come when they are combined with other technologies (LLMs will likely take the credit, but that's okay)